Multilingual Sentiment Analysis Using a Single LSTM Architecture: A Comparative Study on Turkish and English Datasets

Asım Selvi[1*], Emre Satır[2]
[1]Department of Software Engineering, Izmir Katip Celebi University, Izmir, Turkey
2Department of Computer Engineering, Izmir Katip Celebi University, Izmir, Turkey
ORCIDs: 0009-0002-0348-5165, 0000-0002-1950-5549
E-mails: y240237010@ogr.ikcu.edu.tr, emre.satir@ikcu.edu.tr
*Corresponding author.

***Abstract*— Sentiment analysis has become an essential task in Natural Language Processing (NLP), particularly with the growing availability of multilingual textual data. While most studies in the literature focus on monolingual models trained separately for each language, the present study proposes a unified deep learning–based framework that performs sentiment analysis on both Turkish and English texts using a single LSTM architecture. Two datasets were employed: the publicly available IMDb movie review dataset for English and a manually labeled dataset consisting of approximately 2,000 Turkish sentences. Texts in both languages were preprocessed, tokenized, and transformed into fixed-length vector representations through embedding and LSTM layers, and binary sentiment classification was performed using a sigmoid activation function. Experimental results demonstrate that the model achieves high accuracy on the English dataset, benefiting from its large and well-balanced structure, while comparatively lower generalization performance is observed for the Turkish dataset due to its smaller size and limited domain coverage. The findings highlight the importance of dataset scale and linguistic characteristics in multilingual sentiment analysis and show that LSTM-based architecture provides an effective baseline for bilingual sentiment classification. Future work will focus on expanding Turkish data resources and integrating transformer-based multilingual models to improve performance across morphologically rich languages.**

***Keywords*—Multilingual Sentiment Analysis, Natural Language Processing (NLP), Long Short-Term Memory (LSTM), Deep Learning, Cross-Language Modeling**

I.INTRODUCTION

Sentiment analysis is a fundamental research area within the field of Natural Language Processing (NLP), aiming to automatically identify and interpret subjective information, opinions, and emotional expressions in textual data. With the rapid growth of digital communication platforms and user generated content, the volume of available textual data has increased dramatically, creating a strong demand for automated and scalable sentiment analysis techniques.

Previous studies on sentiment analysis have primarily focused on single-language scenarios, where models are trained and evaluated on datasets belonging to a specific language, most commonly English. Such monolingual approaches have achieved notable success in tasks such as opinion classification, emotion detection, and polarity identification. However, these models are generally limited in their ability to generalize across languages and often require language-specific resources, preprocessing pipelines, and training procedures.

More recent research has explored multilingual and cross lingual sentiment analysis, aiming to overcome the limitations of monolingual systems. These studies typically rely on multilingual embeddings, transfer learning, or transformer-based architectures to enable knowledge transfer between languages. While these approaches improve cross-lingual performance, they often focus on model portability or translation-based techniques rather than unified multilingual analysis within a single experimental framework.

In contrast to the existing literature, the present study proposes a unified sentiment analysis framework that simultaneously processes and evaluates Turkish and English texts using a single model architecture. Unlike traditional approaches that train separate models for each language, the proposed method integrates both languages into a shared modeling pipeline and provides comparative statistical outputs across languages. This enables not only sentiment classification but also cross language analytical insights within a consistent experimental setting.

By offering a bilingual sentiment analysis approach that jointly handles structurally different languages such as Turkish and English, this study contributes to the literature by demonstrating the feasibility of multilingual sentiment analysis without relying on language-specific models. This approach supports scalable and extensible sentiment analysis systems and highlights the potential for developing more generalized models capable of operating across multiple linguistic domains.

The remainder of this paper is organized as follows. Section 2 reviews related work in the field of sentiment analysis. Section 3 describes the datasets and the proposed methodology. Section 4 presents the experimental results, and Section 5 discusses the findings and outlines conclusions and future research directions.

II.RELATED WORKS

Sentiment analysis has emerged as a significant research and application area within the field of Natural Language Processing (NLP). It can be performed using a wide range of techniques in artificial intelligence. Within this scope, various artificial intelligence methods and algorithms are employed to detect, extract, and identify emotional content embedded in the semantic structure of texts. Sentiment analysis enables the examination of social media content, news articles, and diverse written texts in an integrated manner with technologies such as text mining and machine learning [21], [25].

Various sentiment analysis techniques aim to generate meaningful data from texts for user evaluations. Among the commonly used methodologies are word clouds, text classification, and association rule mining applications [21]. The analysis of individuals' opinions expressed on virtual platforms has become more accessible and cost-effective compared to face-to-face communication through sentiment analysis applications [25]. Sentiment classifications are generally categorized as positive, negative, and neutral, allowing the extraction of meaningful information from textual data [18].

It is well known that deep learning techniques used in artificial intelligence have significantly contributed to the advancement of sentiment analysis applications. In this context, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architectures are effective in extracting meaningful information from textual data. These models enable the determination of sentiment intensity and the visualization of analytical results [15], [17]. Sentiment analysis studies conducted on Turkish texts confirm the effectiveness of these techniques and contribute to obtaining more accurate results [16].

Sentiment analysis and text mining are not limited to social media platforms but are also applied to legal documents and financial reports. The automated analysis of textual content facilitates the rapid processing of large datasets and contributes to decision-support mechanisms [8]. The integration of machine learning and deep learning algorithms with natural language processing reduces the complexity of analyzing legal texts and alleviates the workload of legal professionals [10]. Sentiment analysis studies conducted in structurally diverse languages such as Turkish encounter various challenges. Structured learning based neural approaches have also been applied to Turkish sentiment analysis on social media data [19]. Language-specific characteristics of Turkish necessitate the customization of techniques and algorithms used in this field [7], [22]. Although these approaches share similarities with universal methods, language-specific adaptations are required for Turkish texts [6], [24].

Data sources play a critical role in the performance of sentiment analysis. Domain-specific multilingual dataset generation methods have been proposed to support cross-lingual sentiment analysis tasks [12]. Data obtained from social media, customer reviews, and product evaluations are analyzed using artificial intelligence applications, providing researchers with in-depth insights [8], [9]. For instance, sentiment analysis has been conducted using customer reviews as input data on ecommerce platforms, resulting in the creation of extensive datasets related to customer experiences [22].

The continuous evolution of sentiment analysis techniques necessitates the regular updating of algorithms and models. Misclassifications occurring on dynamic platforms such as social media may lead to new challenges. Ongoing studies aim to address these issues through more effective algorithms [26]. The use of deep learning models such as BERT introduces significant innovations in analyzing emotional states within texts [9]. Artificial intelligence training processes and the enrichment of datasets continue to advance to enhance model performance. Especially targeted studies conducted for Turkish sentiment analysis have achieved substantial progress in deter mining sentiment polarity within specific categories of terms [24]. These studies provide analytical and efficiency-oriented contributions to the field of natural language processing [26]. Sentiment analysis applications are widely used across various domains, including academic research, commerce, and business environments. Companies optimize their marketing strategies through sentiment analysis of customer reviews and shape product development processes using more reliable data [13]. Through user data obtained from social media, organizations can reach broad customer segments and support decision-making processes [20].

Overall, the literature reflects a rapid methodological evolution in sentiment analysis, driven by advances in model architectures and the availability of large-scale datasets. With the advancement of computational technologies and the emergence of increasingly sophisticated algorithms, the significance of sentiment analysis has continued to grow. The expanding availability of large-scale textual data and the development of advanced machine learning and deep learning techniques have further accelerated progress in this domain. As a result, sentiment analysis has emerged as a crucial enabling technology not only for addressing current analytical needs but also for supporting the development of future intelligent systems.

III.*MATERIALS AND METHODS*

A. Model Architecture and Classification Strategy

In this study, deep learning techniques within the scope of artificial intelligence were employed. Sentiment analysis on Turkish and English texts was performed using a single training model. The texts were transformed into vector representations through Tokenizer and Embedding layers, and an LSTM (Long Short-Term Memory) architecture was adopted. The model utilized a multi-layer LSTM structure and performed binary classification (positive/negative) using a sigmoid activation function in the output layer. This approach aimed to enable users to evaluate output data more effectively and to facilitate decision-support processes. Statistical outputs were generated in percentage form and presented through graphical visualizations.

B. LSTM in Natural Language Processing

The LSTM architecture has provided significant improvements in the analysis of sequential data within the field of Natural Language Processing (NLP). Owing to its capability to learn temporal dependencies, LSTM is widely used in tasks such as text classification, sentiment analysis, and language modeling. LSTM architectures have also demonstrated strong performance in other sequential classification domains beyond text [23]. In a study conducted on Turkish Twitter data, LSTM, CNN-LSTM, and BERT models were compared, and the performance of these algorithms in detecting emotional content in texts was evaluated. The strong memory capacity of LSTM is effective in modeling complex emotional structures in textual data. Dönmez and Becerikli compared not only LSTM but also various deep learning architectures, evaluating the types of data for which each model performs more effectively. The time dependent memory management capabilities of LSTM offer a notable advantage in sentiment analysis tasks [9].

C. Mathematical Formulation and Activation Functions

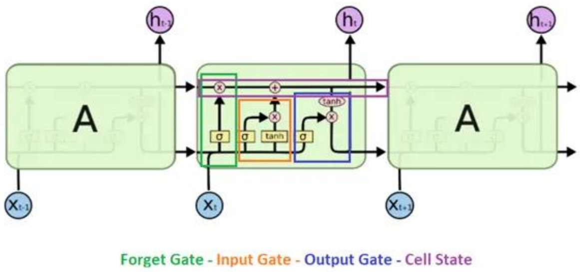

Figure 1: General Structure of the LSTM Architecture

$$i_t = \sigma(W_i[h_{t-1},x_t] + b_i) \quad (1)$$
$$f_t = \sigma(W_f[h_{t-1},x_t] + b_f) \quad (2)$$
$$o_t = \sigma(W_o[h_{t-1},x_t] + b_o) \quad (3)$$
$$C_t = f_t * C_{t-1} + i_t * \tanh(W_C[h_{t-1},x_t]+b_C) \quad (4)$$
$$h_t = o_t * \tanh(C_t) \quad (5)$$

In these equations, σ denotes the logistic sigmoid function, and W represents the weight matrices. The vectors referred to as gates, such as i_t, f_t, and o_t, have the same dimensionality as the hidden state *vector (h)*. The activation functions commonly used in LSTM architectures are the sigmoid and hyperbolic tangent functions [4], [11].

D. LSTM Cell State and Gate Mechanisms)

One of the fundamental components of LSTM is the cell state, which is responsible for information transfer. This structure enables the propagation of meaningful information throughout the network, thereby overcoming the short-term memory problem. To determine which information should be retained or discarded, gates are employed. These gates operate using the sigmoid activation function; information with an output value of 0 is forgotten, while information with an output value of 1 is retained [4], [11].

Forget Gate: This gate determines which information should be forgotten or retained by utilizing information from the previous hidden state (ht−1) and the current input (xt).
Input Gate: This gate is responsible for updating the cell state. The data obtained through sigmoid activation is evaluated together with the hyperbolic tangent (tanh) function to perform the update.
Output Gate: This gate determines the hidden state (ht) to be transferred to the next cell. The information carried within the cell is passed through the tanh activation function and multiplied by the sigmoid output before being propagated [4], [11].

E. Advantages of LSTM for Sentiment Analysis

These mechanisms enable LSTM to learn both short-term and long-term dependencies. Therefore, it achieves high performance in complex text-based tasks such as sentiment analysis.

F. Data Source, Preprocessing, and Transformation Steps

In this study, sentiment analysis was conducted on texts in two languages, Turkish and English. Different data sources were utilized for each language, and the data were prepared for model input using Natural Language Processing (NLP) techniques.

G. English Dataset

For the English texts, the publicly available IMDb Movie Review Dataset [14] provided within the TensorFlow library was used [1]. This dataset consists of 25,000 training samples and 25,000 test samples, each labeled as either positive or negative. Each text is provided in a numerical representation format (token IDs). To ensure compatibility with the model input, the data were processed as follows:

- First, the token ID sequences were converted back into their corresponding word representations,

- Subsequently, the texts were re-tokenized into numerical form using the *Tokenizer*,
- Finally, the sequences were standardized to a maximum length of 200 using the *pad_sequences* function.

H. Turkish Dataset

The Turkish texts consist of an approximately 2,000 sentence dataset created in the Microsoft Excel environment and manually labeled as positive or negative. Prior studies indicate that the choice of preprocessing techniques significantly affects sentiment classification performance [3]. These texts were subjected to the following preprocessing steps:

- Punctuation marks were removed using the expression *re.sub(’[,˙ !?:()"]’, ”, x)*,
- All texts were converted to lowercase,
- Irrelevant words were eliminated using a Turkish stop word list [5],
- Finally, the texts were tokenized at the word level using the *nltk.tokenize.word_tokenize* function [5].

Following these steps, the data were converted into numerical sequences using the *Tokenizer* and aligned to a fixed length of 200 words via the *pad_sequences* function. Earlier semantic representation approaches such as Word2Vec have been widely used to model lexical relations [2].

I. Post-Transformation Preparation

As a result of the preprocessing and transformation procedures applied to both datasets, the data were rendered suitable for the LSTM architecture and fed into the model through the Embedding layer. Each sample was presented to the model as a sequence of embedded vectors with fixed length and dimensionality.

In this manner, texts obtained from different sources were integrated into a unified modeling framework, enabling sentiment analysis to be performed consistently across both languages.

IV.EXPERIMENTAL RESULTS

In the experimental phase of the study, the IMDb English movie review dataset and a custom-prepared Turkish dataset stored in Microsoft Excel format were utilized. All data preprocessing, model training, and evaluation procedures were implemented using the Python programming language and its associated machine learning and deep learning libraries.

The performance of the proposed LSTM-based sentiment analysis model was evaluated using both training and validation datasets. Figure 2 illustrates the sentiment distribution obtained from the model predictions for Turkish and English texts. The results indicate that positive sentiment predictions dominate the Turkish dataset with a rate of 70%, while negative sentiment constitutes 30%. In contrast, the English dataset exhibits a balanced distribution, with 50% positive and 50% negative sentiment classifications. This difference can be attributed to variations in dataset composition and linguistic characteristics.

Figures 3 and 4 present the accuracy and loss curves of the English model, respectively. As observed, the training accuracy increases steadily across epochs, reaching over 92%, while the validation accuracy also shows a consistent upward trend. Concurrently, both training and validation loss values decrease significantly, indicating effective learning and good generalization capability of the model on English texts.

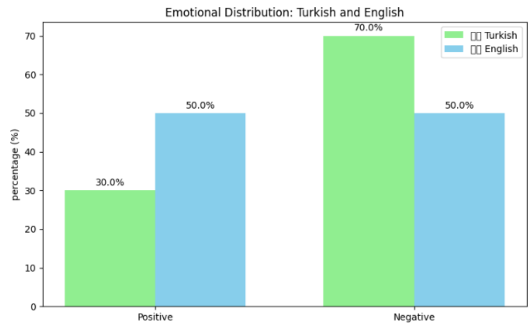

Figure 2: Sentiment distribution percentage for Turkish and English datasets

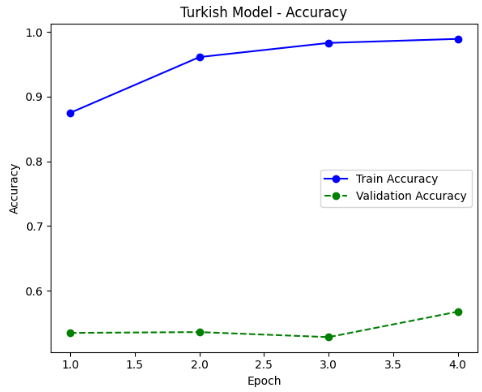

Figure 3: Training and validation accuracy of the English model

Similarly, Figures 5 and 6 depict the performance of the model trained on Turkish data. The training accuracy rapidly approaches 99%, demonstrating the model’s ability to learn sentiment patterns within the dataset. Although fluctuations are observed in validation accuracy and loss values, the overall trend suggests that the model successfully captures sentiment related features in Turkish texts despite the relatively smaller dataset size.

Overall, the experimental results demonstrate that the proposed LSTM-based approach is effective for sentiment analysis in both Turkish and English languages. The model achieves high accuracy values and stable loss convergence, confirming the suitability of deep learning methods for multilingual sentiment analysis tasks.

V. DISCUSSION AND ONCLUSION

The experimental findings indicate that the performance of the model trained using the LSTM architecture varies depending on the size, quality, and linguistic characteristics of the training dataset. In particular, the model trained on the English dataset achieved higher performance levels compared to the Turkish model. This outcome can be attributed to the use of the large-scale and well-annotated IMDb dataset for English, which provides rich linguistic diversity and balanced sentiment distribution.

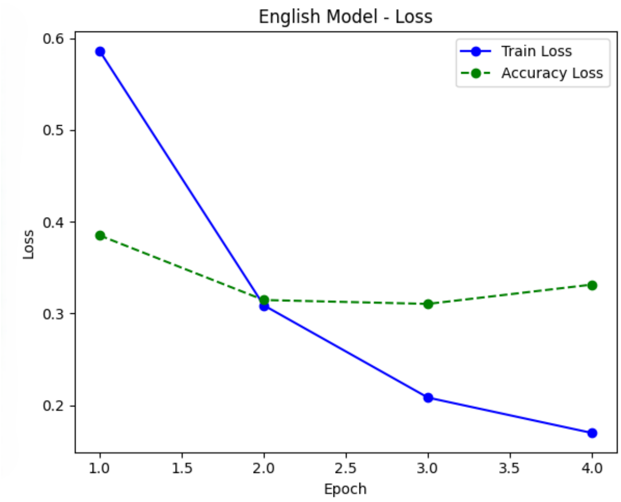

Figure 4: Training and validation loss of the English model

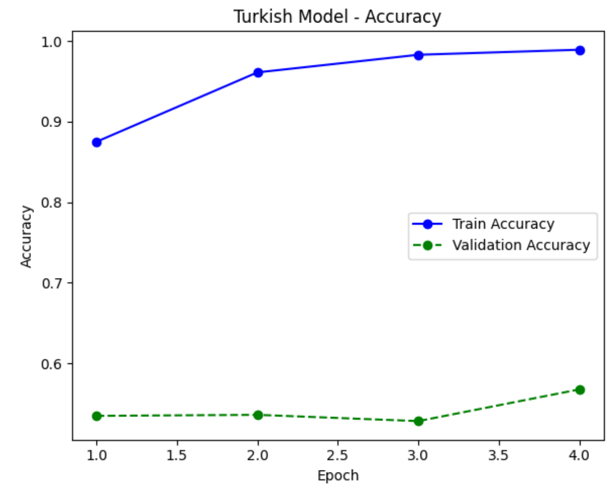

Figure 5: Training and validation accuracy of the Turkish model

In contrast, the Turkish model was trained on a medium scale dataset consisting of approximately 2,000 labeled samples. Due to its relatively limited size and narrower domain coverage, the model exhibited lower generalization performance when compared to the English counterpart. These findings highlight the importance of dataset scale and diversity in achieving robust sentiment classification performance, especially in deep learning-based approaches.

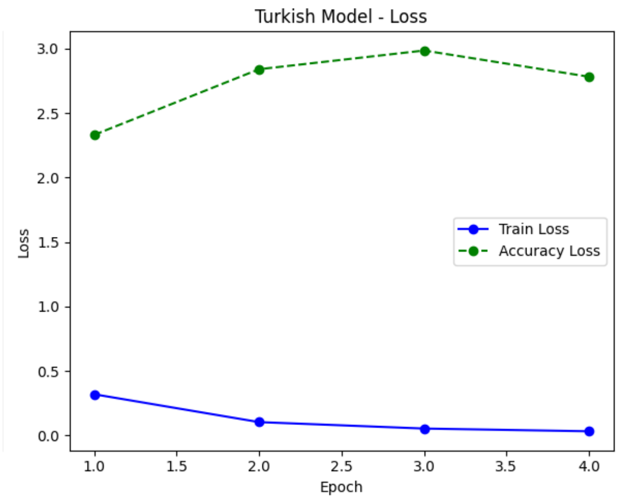

Figure 6: Training and validation loss of the Turkish model

From a methodological perspective, the results confirm that LSTM-based architectures are effective for sentiment analysis tasks in both English and Turkish. However, for morpho logically rich and agglutinative languages such as Turkish, additional linguistic preprocessing, larger datasets, and potentially more advanced architectures may be required to reach comparable performance levels.

Based on these observations, several recommendations can be proposed for future research. First, the expansion of Turkish sentiment datasets in terms of size and domain diversity is essential to improve model robustness and generalization. Second, the integration of transformer-based architectures such as BERT and its multilingual variants may further enhance classification performance, particularly for languages with complex morphological structures. Finally, incorporating domain-specific fine-tuning and data augmentation techniques may help mitigate data scarcity issues and improve sentiment detection accuracy.

Overall, this study demonstrates the feasibility and effectiveness of deep learning-based sentiment analysis in a multilingual context, while also emphasizing the critical role of data quality and model selection in achieving reliable and scalable sentiment analysis systems.

REFERENCES

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. Tensorflow: A system for large-scale machine learning. arXiv preprint arXiv:1605.08695 (2016), https://arxiv.org/abs/1605.08695

[2] Ant, K., Soğukpınar, U., Amasyalı, M.F. Comparison of templates with word2vec in finding semantic relations between words. Journal of Intelligent Systems with Applications 1(1), 13–17 (2018), https://joiswa.com/issue.php?issue=1

[3] Atayolu, Y., Kutlu, Y. Effect of text preprocessing methods on the performance of social media posts classification. Journal of Intelligent Systems with Applications 7(1), 1–6 (2023), https://joiswa.com/indir.php?id=304

[4] Berus, Y., Bente¸sen Yakut, Y. Derin Öğrenme (1d-cnn, rnn, lstm, bilstm) ile enerji tüketim tahmini: Diyarbakır avm Örne˘gi. DÜMF Mühendislik Dergisi (2024). https://doi.org/10.24012/dumf.1415055

[5] Bird, S. Nltk: The natural language toolkit. In: Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions. pp. 69–72. As sociation for Computational Linguistics (2006)

[6] Can, V.B., Şatır, E. Mitigating gender bias in english–spanish machine translation: A parameter-efficient fine-tuning approach. Journal of Intelligent Systems with Applications 8(1), 20–24 (2025)

[7] Demir, Ö., Chawai, A.I.B., Doğan, B. Türkçe metinlerde sözlük tabanli yaklaşimla duygu analizi ve görselleştirme. Int. Periodical of Recent Technologies in Applied Engineering 1(2), 14–22 (2020). https://doi.org/10.35333/porta.2019.98

[8] Demirbilek, M., Demirbilek, S. Google yorumları Üzerinden makine Öğrenme yöntemleri ve amazon comprehend ile duygu analizi. Üniversite Araştırmaları Dergisi 6(4), 452–461 (2023). https://doi.org/10.32329/uad.1383794

[9] Dönmez, H.G., Becerikli, Y. Türkçe twitter verileri Üzerinde duygu analizi: Lstm, cnn-lstm, bert algoritmalarının karşılaştırılması. International Journal of Advanced Natural Sciences and Engineering Re searches 7(6), 132–138 (2023). https://doi.org/10.59287/ijanser.1146

[10] Görentaş, M.B., Uçkan, T., Arlı, N.B. Uyuşmazlık mahkemesi karar larının makine Öğrenmesi yöntemleri ile sınıflandırılması. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi 28(3), 947–961 (2023). https://doi.org/10.53433/yyufbed.1292275

[11] Hochreiter, S., Schmidhuber, J. Long short-term mem ory. Computation 9(8), https://doi.org/10.1162/neco.1997.9.8.1735

[12] İnan, E., Mostafapour, V., Tekbacak, F. Domain-specific evaluation dataset generator for multilingual text analysis. Journal of Intelligent Systems with Applications 2(2), 140–147 (2019), https://joiswa.com/abstract.php?id=84

[13] Karaca, A., Aydın, Ö. Transformatör mimarisi tabanlı derin Öğrenme yöntemi ile türkçe haber metinlerine başlık Üretme. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi 39(1), 485–496 (2023). https://doi.org/10.17341/gazimmfd.963240

[14] Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y., Potts, C. Learning word vectors for sentiment analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. pp. 142–150. Association for Computational Linguistics (2011)

[15] Onan, A. Evrişimli sinir ağı mimarilerine dayalı türkçe duygu analizi. European Journal of Science and Technology pp. 374–380 (2020). https://doi.org/10.31590/ejosat.780609

[16] Onan, A. Türkçe Metin Madenciliği için Dikkat Mekanizması Tabanlı Derin Öğrenme Mimarilerinin Değerlendirilmesi. European Journal of Science and Technology (2022). https://doi.org/10.31590/ejosat.1082379

[17] Onan, A. Türkçe Metin Madenciliği için Çalışan Bellek Bağlantıları Tabanlı Uzun Kısa Süreli Bellek Mimarisi. European Journal of Science and Technology (2022). https://doi.org/10.31590/ejosat.1080239

[18] Tokcaer, S. Türkçe metinlerde duygu analizi. Journal of Yaşar University 16(63), 1516–1536 (2021). https://doi.org/10.19168/jyasar.928843

[19] Ülgen, O., Öğrenci, A.S. Structured learning based turkish sentiment analysis. Journal of Intelligent Systems with Applications 2(2), 90–93 (2019), https://joiswa.com/abstract.php?id=71

[20] Uyrun, Ö.F., Sabuncu, İ. Sosyal Medya ve Diğer Yatırım Aracı Verilerine Dayalı Hisse Senedi Değeri Tahmini. Acta Infologica 5(2), 267–285 (2021). https://doi.org/10.26650/acin.934130

[21] Yürütücü, Y. Ön eğitimli dil modelleriyle duygu analizi. İstanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi 5(1), 46 53 (2023). https://doi.org/10.47769/izufbed.1312032

[22] Zada, A.J.J., Albayrak, A. Duygu analizi ve topluluk Öğrenmesi yaklaşımları ile kullanıcı yorumlarının analizi. Düzce Üniver sitesi Bilim Ve Teknoloji Dergisi 11(4), 1725–1732 (2023). https://doi.org/10.29130/dubited.1102181

[23] Çancıoğlu, E., ¸Sahin, S., İşler, Y. Heart sounds analysis and classification based on long-short term memory.

Journal of Intelligent Systems with Applications 3(1), 25–28 (2020), https://joiswa.com/issue.php?issue=5

[24] Çetin, F.S., Eryiğit, G. Türkçe hedef tabanlı duygu analizi ˙ Için alt görevlerin ˙ Incelenmesi– hedef terim, hedef kategori ve duygu sınıfı belirleme. Bili¸sim Teknolojileri Dergisi 11(1), 43–56 (2018). https://doi.org/10.17671/gazibtd.325865

[25] Öğe, B.C., Kayaalp, F. Farklı sınıflandırma algoritmaları ve metin temsil yöntemlerinin duygu analizinde performans kar¸sıla¸stırılması. Düzce Üniversitesi Bilim ve Teknoloji Dergisi 9(6), 406–416 (2021). https://doi.org/10.29130/dubited.1015320

[26] Şahınaslan, Ö, Dalyan, H., Şahınaslan, E. Naive Bayes Sınıflandırıcısı Kullanılarak Youtube Verileri Üzerinden Çok Dilli Duygu Analizi. Bilişim Teknolojileri Dergisi 15(2), 221–229 (2022). https://doi.org/10.17671/gazibtd.999960